

Cognitive Hacking and the Value of Information

George Cybenko, Annarita Giani, Paul Thompson
Thayer School of Engineering and
Institute for Security Technology Studies
Dartmouth College
Hanover NH 03755

1.0 Background

Computer and network security present great challenges to our evolving information society and economy. The variety and complexity of cybersecurity attacks that have been developed parallel the variety and complexity of the information technologies that have been deployed, with no end in sight for either. We delineate between two classes of information systems attacks: *autonomous* attacks and *cognitive* attacks. Autonomous attacks operate totally within the fabric of the computing and networking infrastructures. For example, the well-know Unicode attack against older, unpatched versions of Microsoft's Internet Information Server (IIS) can lead to root/administrator access. Once such access is obtained, any number of undesired activities by the attacker is possible. For example, files containing private information such as credit card numbers can be downloaded and used by an attacker. Such an attack does not require any intervention by users of the attacked system, hence we call it an "autonomous" attack. By contrast, a *cognitive* attack requires some change in users' behavior, affected by manipulating their perception of reality. The attack's desired outcome cannot be achieved unless human users change their behaviors in some way. Users' modified actions are a critical link in a cognitive attack's sequencing.

Cognitive attacks can be overt or covert. No attempt is made to conceal overt cognitive attacks, e. g., website defacements. Provision of misinformation, the intentional distribution or insertion of false or misleading information intended to influence reader's decisions and / or activities, is covert cognitive hacking. The Internet's open nature makes it an ideal arena for dissemination of misinformation. Cognitive hacking differs from social engineering, which, in the computer domain, involves a hacker's psychological tricking of legitimate computer system users to gain information, e.g., passwords, in order to launch an autonomous attack on the system.

2.0 Value of Information – Information Theoretic and Economic Models

Information theory has been used to analyze the value of information in horse races and in optimal portfolio strategies for the stock market [2]. We have begun to investigate the applicability of this analysis to cognitive hacking. So far we have considered the simplest case, that of a horse race with two horses.

2.1 An Information Theoretic Model of Cognitive Hacking

Sophisticated hackers can use information theoretic models of a system to define a gain function and conduct a sensitivity analysis of its parameters. The idea is to identify and target the most sensitive variables of the system, since even slight alterations of their value may influence people's behavior. For example, specific information on the health of a company might help stock brokers predict fluctuations in the value of its shares. A cognitive hacker manipulates the victim's perception of the likelihood of winning a high payoff in a game. Once the victim has decided to play, the cognitive hacker influences which strategy the victim chooses.

2.1.1 A Horse Race

Here is a simple model illustrating this kind of exploit. A horse race is a system defined by the following elements [2]

- There are m horses running in a race
- each horse i is assigned a probability p_i of winning the race (so $\{p_i\}_{i=1..m}$ is a probability distribution)
- each horse i is assigned an odds o_i signifying that a gambler that bet b_i dollars on horse i would win $b_i o_i$ dollars in case of victory (and suffer a total loss in case of defeat).

If we consider a sequence of n independent races, it can be shown that the average rate of the wealth gained at each race is given by

$$W(b, p, o) = \sum_{i=1}^m p_i \log b_i o_i$$

where b_i is the percentage of the available wealth invested on horse i at each race. So the betting strategy that maximizes the total wealth gained is obtained by solving the following optimization problem

$$W(p, o) = \max_b W(b, p, o) = \max_b \sum_{i=1}^m p_i \log b_i o_i$$

subject to the constraint that the b_i 's add up to 1. It can be shown that this solution turns

out to be simply $b=p$ (proportional betting) and so $W(p, o) = \sum_{i=1}^m p_i \log p_i o_i$.

Thus, a hacker can predict the strategy of a systematic gambler and make an attack with the goal of deluding the gambler on his/her future gains. For example, a hacker might lure an indecisive gambler to invest money on false prospects. In this case it would be useful to understand how sensitive the function W is to p and o and tamper with the data in order to convince a gambler that it is worth playing (because W appears illusionary larger than it actually is).

To study the sensitivity of W to its domain variables we consider the partial derivatives of W with respect to p_i and o_i and see where they assume the highest values. This gives us information on how steep the function W is on subsets of its domain.

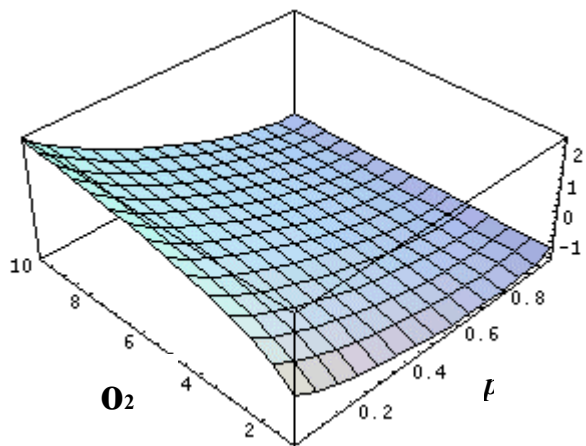
If we consider the special case of races involving only two horses ($m=2$), then we have

$$W(p, o_1, o_2) = p \log p o_1 + (1 - p) \log(1 - p) o_2$$

- $\frac{\partial W}{\partial p}(p, o_1, o_2) = \log\left(\frac{p}{1-p} \frac{o_1}{o_2}\right)$
- $\frac{\partial W}{\partial o_1}(p, o_1, o_2) = \frac{p}{o_1}$
- $\frac{\partial W}{\partial o_2}(p, o_1, o_2) = \frac{1-p}{o_2}$

Thus, if we fix one of the variables then we can conduct a graphic analysis of those functions with a 3D plot.

Case 1. o_1 is constant.



$$W(p, 3, o_2)$$

This is the doubling rate function. The most sensitive parameter to let W increase is o_2 . Increasing this variable W grows at a fast rate for low values of p and grows with a smaller rate for higher values of p .

2.1.2 Applying the Horse Race Example to the Internet

Misinformation was used by Mark Jakob in a cognitive attack [5]. Jakob posted a bogus release regarding the company Emulex on Internet Wire, a Los Angeles press-release distribution firm. The release was picked up by several business news services and widely redistributed without independent verification. Jakob sold Emulex short and profited, while other investors lost large sums of money selling the stock as its value fell sharply in response to the misinformation.

In this example the two horses are: horse 1, Emulex stock goes up; and horse 2, Emulex stock goes down. First the cognitive hacker makes the victim want to play the game by making the victim think that he can make a large profit through Emulex stock transactions. This is done by spreading misinformation about Emulex, whether positive or negative, but news that, if true would likely cause the stock's value to either sharply

increase, or decrease, respectively. Positive misinformation might be the news that Emulex had just been granted a patent that could lead to a cure for AIDS. Negative misinformation might be that Emulex was being investigated by the Securities and Exchange Commission (SEC) and that the company was forced to restate 1998 and 1999 earnings. This fraudulent negative information was in fact posted by Jakob.

2.2 Theories of the Firm and Cognitive Hacking

Much attention in economics has been devoted to theories of the market. The economic actor has been modeled as enjoying perfect, costless information. Such analyses, however, are not adequate to explain the operation of firms. Theories of the firm provide a complementary economic analysis taking into account transaction and organizing costs, hierarchies, and other factors left out of idealized market models. It has been argued that information technology will transform the firm, such that “. . . the fundamental building blocks of the new economy will one day be ‘virtual firms’, ever-changing networks of subcontractors and freelancers, managed by a core of people with a good idea” [3]. Others argue that more efficient information flows not only lower transaction costs, thereby encouraging more outsourcing, but also lower organization costs, thereby encouraging the growth of larger companies [1]. More efficient information flow implies a more standardized, automated processing of information, which is susceptible to cognitive attack. Schneier [6] attributes the earliest conceptualization of computer system attacks as physical, syntactic, and semantic to Martin Libicki, who describes semantic attacks in terms of misinformation being inserted into interactions among intelligent software agents [4]. Libicki was describing information warfare, but semantic, or cognitive, attacks can be directed against business systems, as well.

3.0 Cognitive Hacking Countermeasures

Cognitive hacking on the internet is an evolving and growing activity, often criminal and prosecutable. Technologies for preventing, detecting and prosecuting cognitive hacking are still in their infancies. Given the variety of approaches to and the very nature of cognitive hacking, *preventing* cognitive hacking reduces either to preventing unauthorized access to information assets (such as in web defacements) in the first place or detecting posted misinformation before user behavior is affected (that is, before behavior is changed but possibly after the misinformation has been disseminated). The latter may not involve unauthorized access to information, as for instance in "pump and dump" schemes that use newsgroups and chat rooms. By definition, *detecting* a successful cognitive hack would involve detecting that the user behavior has already been changed. We are not considering detection in that sense at this time.

3.1 Single Source Cognitive Hacking

In this section, we develop a few possible approaches for the single source problem. By single source, we mean situations in which redundant, independent sources of information about the same topic are not available. An authoritative corporate personnel database would be an example. The multiple source problem will not be discussed here.

3.1.1 Authentication of source

This technique involves due diligence in authenticating the information source and ascertaining its reliability. Various relatively mature certification and PKI technologies can be used to detect spoofing of an information server. Additionally, reliability metrics can be established for an information server or service by scoring its accuracy over repeated trials and different users.

3.1.2 Information "trajectory" modeling

This approach requires building a model of a source based on statistical historical data or some sort of analytic understanding of how the information relates to the real world. For example, weather data coming from a single source (website or environmental sensor) could be calibrated against historical database (from previous years) or predictive model (extrapolating from previous measurements). A large deviation would give reason for hesitation before committing to a behavior or response.

4.0 Summary

Cognitive hacking represents a new kind of threat and requires new kinds of tools for preventing it. This paper has defined the basic concepts and presented a simple information theoretic model of cognitive attacks. Approaches for preventing single source cognitive hacking have been proposed, as well.

5.0 References

1. Agre, Philip. 2001. "The market logic of information" *Knowledge, Technology, and Policy* vol. 13, no. 1, p. 67-77.
2. Cover, Thomas A. and Thomas, Joy A. 1991. *Elements of Information Theory* New York: Wiley
3. *Economist*. 2002. "Re-engineering in real time" 31 January. http://www.economist.com/surveys/PrinterFriendly.cfm?Story_ID=949093.
4. Libicki, Martin. 1994. "The mesh and the Net: Speculations on armed conflict in an age of free silicon" National Defense University McNair Paper 28 <http://www.ndu.edu/ndu/inss/macnair/mcnair28/m028cont.html>.
5. Mann, Bill. 2000. "Emulex fraud hurts all" *The Motley Fool*. <http://www.fool.com/news/foolplate/2000/foolplate000828.htm>
6. Schneier, Bruce. 2000. "Semantic attacks: The third wave of network attacks" *Crypto-gram Newsletter* October 15, 2000. <http://www.counterpane.com/crypto-gram-0010.html>.